

Recovery of Signals with Low Density

Christoph Studer

Abstract

Sparse signals (i.e., vectors with a small number of non-zero entries) build the foundation of most kernel (or nullspace) results, uncertainty relations, and recovery guarantees in the sparse signal-processing and compressive-sensing literature. In this paper, we introduce a novel *signal-density measure* that extends the common notion of sparsity to non-sparse signals whose entries' magnitudes decay rapidly. By taking into account such magnitude information, we derive a more general and less restrictive kernel result and uncertainty relation. Furthermore, we demonstrate the effectiveness of the proposed signal-density measure by showing that orthogonal matching pursuit (OMP) provably recovers low-density signals with up to $2\times$ more non-zero coefficients compared to that guaranteed by standard results for sparse signals.

I. INTRODUCTION

In this paper, we consider kernel (or nullspace) results providing conditions for which $\mathbf{Ax} \neq \mathbf{0}_{M \times 1}$ and uncertainty relations for pairs of signals (vectors) $\mathbf{x} \in \mathbb{C}^{N_a}$ and $\mathbf{z} \in \mathbb{C}^{N_b}$ that satisfy $\mathbf{Ax} = \mathbf{Bz}$, where $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ and $\mathbf{B} \in \mathbb{C}^{M \times N_b}$ are dictionaries, i.e., matrices whose columns are normalized to unit ℓ_2 -norm. We furthermore study conditions for which orthogonal matching pursuit (OMP) recovers signals \mathbf{x} from an underdetermined system of linear equations $\mathbf{y} = \mathbf{Ax}$ with $M < N_a$.

A. Contributions and Paper Outline

In contrast to existing kernel results, uncertainty relations, and recovery guarantees that have been derived for *sparse* signals, i.e., for vectors with only a small number of nonzero entries, [1]–[11], the results developed in this paper make use of the following definition (see Section II for the details).

Definition 1 (δ -Density): For a non-zero signal $\mathbf{x} \in \mathbb{C}^{N_a}$, we define its δ -density as follows:

$$\delta(\mathbf{x}) = \|\mathbf{x}\|_1 / \|\mathbf{x}\|_\infty. \quad (1)$$

C. Studer is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY (e-mail: studer@cornell.edu). Website: <http://www.csl.cornell.edu/~studer/>

For an all-zero signal $\mathbf{x} = \mathbf{0}_{N_a \times 1}$, we define $\delta(\mathbf{x}) = 0$.

Our results in Section III show that if \mathbf{A} has incoherent columns and the signal \mathbf{x} has sufficiently low δ -density, then $\mathbf{Ax} \neq \mathbf{0}_{M \times 1}$. Furthermore, if the columns among \mathbf{A} and \mathbf{B} are incoherent, then two signals \mathbf{x} and \mathbf{z} satisfying $\mathbf{Ax} = \mathbf{Bz}$ cannot have low density at the same time. In Section IV, we show that orthogonal matching pursuit (OMP) is able to recover up to $1/\mu_a$ non-zero entries of a signal \mathbf{x} from $\mathbf{y} = \mathbf{Ax}$ given the coefficients have low density and decay sufficiently fast; here $\mu_a = \max_{i \neq j} |\mathbf{a}_i^H \mathbf{a}_j|$ is the coherence parameter of \mathbf{A} . This result not only improves upon standard recovery guarantees for OMP by up to $2\times$, but also provides a practical way of recovering the dominant coefficients (in magnitude) of signals with low density. We conclude in Section V. All proofs are relegated to the appendices.

B. Notation

Lowercase and uppercase boldface letters stand for column vectors and matrices, respectively. For the matrix \mathbf{M} , we denote its adjoint and (left) pseudo-inverse by \mathbf{M}^H and $\mathbf{M}^\dagger = (\mathbf{M}^H \mathbf{M})^{-1} \mathbf{M}^H$, respectively; \mathbf{I}_M and $\mathbf{0}_{M \times N}$ denotes the $M \times M$ identity and $M \times N$ all-zero matrix, respectively. The i -th column of the matrix \mathbf{M} is denoted by \mathbf{m}_i . Sets are designated by upper-case calligraphic letters; the cardinality of the set \mathcal{S} is $|\mathcal{S}|$. The matrix $\mathbf{M}_{\mathcal{S}}$ is obtained from \mathbf{M} by retaining the columns of \mathbf{M} with indices in \mathcal{S} ; the vector $\mathbf{v}_{\mathcal{S}}$ is obtained analogously from \mathbf{v} . We define $[x]^+ = \max\{x, 0\}$ for $x \in \mathbb{R}$.

II. PROPERTIES OF THE δ -DENSITY

The δ -density in Definition 1 enables a finer characterization of important signal properties than the signal sparsity $\|\mathbf{x}\|_0$, which simply counts number of non-zero entries in \mathbf{x} . We next summarize the key properties of the proposed density measure.

The following result establishes an upper bound on $\delta(\mathbf{x})$; a short proof is given in Appendix A-A.

Lemma 1 (δ -Density vs. Signal Sparsity): For signals $\mathbf{x} \in \mathbb{C}^{N_a}$, the δ -density in Definition 1 satisfies

$$0 \leq \delta(\mathbf{x}) \leq \|\mathbf{x}\|_0 \leq N_a. \quad (2)$$

Equality $\delta(\mathbf{x}) = \|\mathbf{x}\|_0$ is established if and only if the non-zero entries of \mathbf{x} have constant modulus.

The result in (2) implies that if the signal \mathbf{x} is sparse, i.e., $\|\mathbf{x}\|_0$ is smaller than its ambient dimension N_a , then it must also have low density (bounded by $\|\mathbf{x}\|_0$). Furthermore, for signals whose non-zero entries are constant modulus, both the δ -density $\delta(\mathbf{x})$ and the signal sparsity $\|\mathbf{x}\|_0$ coincide. In contrast, signals with low δ -density must not necessarily be sparse. As an example, consider the following signal class.

Definition 2 (α -Decaying Signal): We define an α -decaying signal as a vector $\mathbf{x} \in \mathbb{R}^{N_a}$ whose entries satisfy $x_i = \alpha^{i-1}$, $i = 1, \dots, N_a$ with $0 < \alpha < 1 - 1/N_a$.

For such signals, the signal sparsity equals the ambient dimension N_a as all entries are non-zero; in contrast, the δ -density can be bounded by

$$\delta(\mathbf{x}) \leq \frac{1}{1 - \alpha} < \|\mathbf{x}\|_0 = N_a.$$

Hence, the δ -density does not simply count the number of nonzero entries, but also captures crucial magnitude information of the signal's non-zero coefficients—all our results developed in Sections III and IV make use of this particular property.

We now show that the δ -density not only exhibits similarity to the signal sparsity $\|\mathbf{x}\|_0$, but also some fundamental differences. The following result is a consequence of Definition 1.

Lemma 2 (Inhomogeneity): The δ -density $\delta(\mathbf{x})$ and $\|\mathbf{x}\|_0$ are invariant to positive scalings.

As a consequence, the δ -density and the sparsity $\|\mathbf{x}\|_0$ are not norms. In contrast to the sparsity $\|\mathbf{x}\|_0$, the δ -density also violates the triangle inequality. A short proof is given in Appendix A-B.

Lemma 3 (Triangle Inequality): In general, the δ -density does not satisfy the triangle inequality.

This result has negative consequences for uniqueness proofs that rely on the δ -density. Specifically, a common way of establishing uniqueness for sparse signals (see, e.g., [8], [9], [11]) is to consider two signals \mathbf{x} and \mathbf{x}' , and investigate $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}'$, which implies $\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{0}_{M \times 1}$. Thanks to the triangle inequality, the sparsity $\|\mathbf{x} - \mathbf{x}'\|_0$ can now be bounded by $\|\mathbf{x}\|_0 + \|\mathbf{x}'\|_0$; this approach, however does not apply to the δ -density, which inhibits the derivation of similar uniqueness guarantees. We therefore provide an alternative recovery condition in Section IV for OMP.

III. KERNEL RESULT AND UNCERTAINTY RELATION

We now develop a novel kernel (or nullspace) result and uncertainty relation for signals with low δ -density. All our results make use of the following definitions.

Definition 3 (Dictionary Coherence): Let $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ be a dictionary. Then $\mu_a = \max_{i \neq j} |\mathbf{a}_i^H \mathbf{a}_j|$ is the coherence of the dictionary \mathbf{A} . The coherence μ_b of a dictionary $\mathbf{B} \in \mathbb{C}^{M \times N_b}$ is defined analogously.

Definition 4 (Mutual Coherence): Let $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ and $\mathbf{B} \in \mathbb{C}^{M \times N_b}$ be two dictionaries. Then $\mu_m = \max_{i,j} |\mathbf{a}_i^H \mathbf{b}_j|$ is the mutual coherence between the dictionaries \mathbf{A} and \mathbf{B} .

Note that for an orthonormal basis (ONB) $\mathbf{A} \in \mathbb{C}^{M \times M}$, we have $\mu_a = 0$. For a pair of ONBs $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{M \times M}$, we have the following lower bound: $\mu_m \geq 1/\sqrt{M}$ (see, e.g., [1], [7]).

A. Kernel Result

The next lemma is key for establishing our δ -density-based kernel result; the proof is given in Appendix B-A.

Lemma 4 (Bounds on the ℓ_∞ Matrix Norm): Let $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ be a dictionary with coherence μ_a and $\mathbf{x} \in \mathbb{C}^{N_a}$ a nonzero signal. Then, the following inequalities hold:

$$1 - \mu_a(\delta(\mathbf{x}) - 1) \leq \frac{\|\mathbf{A}^H \mathbf{A} \mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} \leq 1 + \mu_a(\delta(\mathbf{x}) - 1). \quad (3)$$

This result resembles lower and upper bounds on the ℓ_∞ matrix norm, which corresponds to the maximum absolute row sum of the Gram matrix $\mathbf{A}^H \mathbf{A}$. The bounds in (3), however, depend on the δ -density of the signal \mathbf{x} and enable us to establish the following kernel (or nullspace) result.

Theorem 5 (Kernel Result): Let $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ be a dictionary with non-trivial coherence $\mu_a \neq 0$ and $\mathbf{x} \in \mathbb{C}^{N_a}$ be a non-zero signal. If

$$\delta(\mathbf{x}) < 1 + 1/\mu_a, \quad (4)$$

then \mathbf{x} cannot be in the kernel of \mathbf{A} , i.e., $\mathbf{A} \mathbf{x} \neq \mathbf{0}_{M \times 1}$.

Proof: Assume that $\mathbf{x} \neq \mathbf{0}_{N_a \times 1}$ and $\mathbf{x} \in \ker(\mathbf{A})$. From the left-hand side (LHS) of (3), it follows that $0 \geq (1 - \mu_a(\delta(\mathbf{x}) - 1))\|\mathbf{x}\|_\infty$ and hence, we have $\delta(\mathbf{x}) \geq 1 + 1/\mu_a$. As a consequence, if $\delta(\mathbf{x}) < 1 + 1/\mu_a$, then $\mathbf{x} \notin \ker(\mathbf{A})$. ■

This result implies that non-sparse signals (e.g., with up to N_a non-zero entries) having sufficiently low density *cannot* be in the kernel of an incoherent dictionary. To see this, consider an α -decaying signal with $0 < \alpha < 1 - (1 + 1/\mu_a)^{-1}$. For such signals, we have $\|\mathbf{x}\|_0 = N_a$, whereas the δ -density based condition in (4) is always met. Hence, suitably defined density measures enable us to establish the important fact that not only sufficiently sparse signals cannot be in the kernel of incoherent dictionaries but also certain non-sparse signals with sufficiently low δ -density. We note that Theorem 5 recovers existing nullspace conditions as a special case. From $\delta(\mathbf{x}) \leq \|\mathbf{x}\|_0$ we obtain the well-known (and more restrictive) condition $\|\mathbf{x}\|_0 < 1 + 1/\mu_a$ (see, e.g., [1], [5], [11]).

B. Uncertainty Relation

We next provide a δ -density-based uncertainty relation for pairs of dictionaries; the proof is given in Appendix B-B.

Theorem 6 (Uncertainty Relation): Let $\mathbf{x} \in \mathbb{C}^{N_a}$ and $\mathbf{z} \in \mathbb{C}^{N_b}$ be non-zero vectors satisfying $\mathbf{A} \mathbf{x} =$

$\mathbf{B}\mathbf{z}$, and let $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ and $\mathbf{B} \in \mathbb{C}^{M \times N_b}$ be two dictionaries. Then, the following inequality holds

$$[1 - \mu_a(\delta(\mathbf{x}) - 1)]^+ [1 - \mu_b(\delta(\mathbf{z}) - 1)]^+ \leq \delta(\mathbf{x})\delta(\mathbf{z})\mu_m^2. \quad (5)$$

This uncertainty relation generalizes the result in [8], [9] for sparse signals to the case of signals characterized by the δ -density. As a consequence of Theorem 6, we have the following result.

Corollary 7: For a pair of ONBs \mathbf{A} and \mathbf{B} , the following uncertainty relation holds: $1/\mu_m^2 \leq \delta(\mathbf{x})\delta(\mathbf{y})$.

For maximally-incoherent ONBs (e.g., \mathbf{A} being a Hadamard basis and \mathbf{B} the identity), we have $\mu_m = 1/\sqrt{M}$ [1]. In this case, the uncertainty relation in Corollary 7 leads to $M \leq \delta(\mathbf{x})\delta(\mathbf{y})$, which generalizes the well-known uncertainty relation $M \leq \|\mathbf{x}\|_0 \|\mathbf{y}\|_0$ by Donoho and Stark [12]. As a consequence, for a pair of incoherent ONBs, a signal with low density in basis \mathbf{A} cannot have low-density in basis \mathbf{B} , and vice versa. Again, we emphasize that sparsity is not the key property that is required to establish such uncertainty relations, but rather a suitably-chosen measure of signal density.

IV. RECOVERY OF SIGNALS WITH LOW δ -DENSITY

We next show that signals having sufficiently low density and a fast decaying magnitude profile can be recovered perfectly via OMP. Our condition guarantees recovery of such signals with up to $2\times$ more non-zero coefficients compared to existing, sparsity-based recovery guarantees.

A. Orthogonal Matching Pursuit (OMP)

We now briefly review the OMP algorithm [11], [13], [14]. OMP is an iterative method used to recover sparse vectors \mathbf{x} from $\mathbf{y} = \mathbf{A}\mathbf{x}$. After initializing the residual $\mathbf{r}^{(0)} = \mathbf{y}$ and an empty support set $\mathcal{S}^{(0)} = \emptyset$, OMP selects a column of the dictionary \mathbf{A} in every iteration $t = 1, \dots, t_{\max}$ according to

$$\hat{k}^{(t)} = \arg \max_{i \in \mathcal{R}^{(t-1)}} \left| \mathbf{a}_i^H \mathbf{r}^{(t-1)} \right|. \quad (6)$$

Here, the set $\mathcal{R}^{(t-1)} = \{1, \dots, N_a\} \setminus \mathcal{S}^{(t-1)}$ contains all remaining indices that are not (yet) in the support set $\mathcal{S}^{(t-1)}$. Then, the index $\hat{k}^{(t)}$ in (6) is added to the new support set $\mathcal{S}^{(t)} = \mathcal{S}^{(t-1)} \cup \hat{k}^{(t)}$ and a new residual is computed as

$$\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{A}_{\mathcal{S}^{(t)}} \hat{\mathbf{x}}_{\mathcal{S}^{(t)}} = (\mathbf{I}_M - \mathbf{A}_{\mathcal{S}^{(t)}} \mathbf{A}_{\mathcal{S}^{(t)}}^\dagger) \mathbf{y}, \quad (7)$$

where $\hat{\mathbf{x}}_{\mathcal{S}^{(t)}}$ is the least-squares estimate of the signal's coefficients on the current support set $\mathcal{S}^{(t)}$. The above iterative procedure is repeated until a (predefined) number of iterations t_{\max} is reached. The algorithm's outputs are the least-squares estimate $\hat{\mathbf{x}}_{\mathcal{S}^{(t_{\max})}}$ and the support-set estimate $\mathcal{S}^{(t_{\max})}$.

B. Recovery Guarantee

We now detail our δ -density-based condition for which OMP is guaranteed to identify the indices associated to the largest t_{\max} coefficients in the signal \mathbf{x} from $\mathbf{y} = \mathbf{A}\mathbf{x}$. Furthermore, if the signal \mathbf{x} has exactly t_{\max} non-zero entries (i.e., $\|\mathbf{x}\|_0 = t_{\max}$), then OMP will recover the signal \mathbf{x} perfectly. Concretely, we have the following recovery guarantee; the proof is given in Appendix C.

Theorem 8 (δ -Density-Based Recovery Guarantee): Assume a non-trivial coherence $\mu_a \neq 0$. If the maximum number of OMP iterations satisfies

$$t_{\max} < 1 + 1/\mu_a \quad (8)$$

and, in every iteration $t = 1, \dots, t_{\max}$, the coefficients of the vector \mathbf{x} to be recovered satisfy

$$\delta(\mathbf{x}_{\mathcal{R}^{(t-1)}}) < \frac{1}{2}(1 + 1/\mu_a - (t - 1)), \quad (9)$$

then OMP is guaranteed to select an atom associated to (one of) the largest coefficient(s) from the vector $\mathbf{x}_{\mathcal{R}^{(t)}}$ in iteration t . Furthermore, the support set $\mathcal{S}^{(t)}$ contains the indices associated to the $t = |\mathcal{S}^{(t)}|$ largest (in magnitude) entries in \mathbf{x} .

The first condition (8) ensures that the pseudo-inverse $\mathbf{A}_{\mathcal{S}^{(t)}}^\dagger$ in (7) exists in every iteration. The second condition (9) imposes a constraint on the δ -density of the signal to be recovered—more precisely, it imposes conditions on the magnitude decay of the non-zero entries in \mathbf{x} .

C. Discussion of Theorem 8

To make Theorem 8 more explicit, consider the first iteration of OMP, where we have the empty set $\mathcal{S}^{(0)} = \emptyset$ and $\mathcal{R}^{(0)} = \{1, \dots, N_a\}$. In this situation, condition (9) requires

$$\delta(\mathbf{x}) < \frac{1}{2}(1 + 1/\mu_a) \quad (10)$$

to ensure that OMP selects the column associated with a largest entry in \mathbf{x} .¹ In words, if the δ -density of the signal \mathbf{x} to be recovered is below the RHS of (10), OMP will identify the index of the largest (in magnitude) entry in \mathbf{x} .

In the second iteration (for $t = 2$), condition (9) requires

$$\delta(\mathbf{x}_{\mathcal{R}^{(1)}}) < \frac{1}{2}(1 + 1/\mu_a - 1) \quad (11)$$

¹There could, in general, be multiple entries with the same magnitude; OMP is guaranteed to select one of them.

where $\mathcal{R}^{(1)}$ contains all indices $\{1, \dots, N_a\}$ except the one selected in the first iteration. While the RHS in (11) is more restrictive than (10), the δ -density $\delta(\mathbf{x}_{\mathcal{R}^{(1)}})$ is now potentially smaller too—this, however, depends on the decay profile of the magnitudes in \mathbf{x} . Rewriting (9) reveals that

$$\delta(\mathbf{x}_{\mathcal{R}^{(t-1)}}) < C - \frac{1}{2}(t-1) \quad (12)$$

with $C = \frac{1}{2}(1 + 1/\mu_a)$, ensures that OMP selects the correct atoms in every iteration t (given (8) is satisfied). Hence, by successively removing the t largest entries in \mathbf{x} , we require the δ -density of $\mathbf{x}_{\mathcal{R}^{(t-1)}}$ to decay at least by $C - \frac{1}{2}(t-1)$.

We emphasize that for signals whose non-zero entries have constant modulus, Theorem 8 collapses to the well-known OMP recovery threshold [3], [11]

$$\|\mathbf{x}\|_0 < \frac{1}{2}(1 + 1/\mu_a). \quad (13)$$

For such signals, we have $\delta(\mathbf{x}_{\mathcal{R}^{(t-1)}}) = \|\mathbf{x}\|_0 - (t-1)$, which implies that the condition (12) is met in all OMP iterations (by setting $t_{\max} = \|\mathbf{x}\|_0$). This observation implies (i) that our condition does not contradict existing recovery guarantees for OMP and (ii) that sparse signals with constant modulus are worst-case signals from a Theorem 8 and OMP viewpoint. We note that property (ii) is a rather well-known fact that can also be observed via numerical simulations.

We finally show a simple example that demonstrates superiority of Theorem 8 over the standard recovery guarantee (13). Consider again an α -decaying signal as in Definition 2 and recall that for such signals $\delta(\mathbf{x}) \leq (1 - \alpha)^{-1}$. If we remove the largest t entries in \mathbf{x} , the bound $\delta(\mathbf{x}_{\mathcal{R}^{(t)}}) \leq (1 - \alpha)^{-1}$ continues to hold. Combining this bound with (9), we obtain

$$t < 2 + 1/\mu_a - 2(1 - \alpha)^{-1}.$$

By letting $\alpha \rightarrow 0$, corresponding to very fast coefficient decay, we see that one can perform $t_{\max} < 1/\mu_a$ OMP iterations, without violating the conditions (8) and (9) of Theorem 8. Hence, OMP is able to identify the indices associated to the largest t_{\max} entries (in magnitude), even for non-sparse signals having up to N_a non-zero entries. By truncating an α -decaying signal to the largest $t_{\max} < 1/\mu_a$ entries, OMP is able to *perfectly* recover the signal; this is roughly $2\times$ less restrictive than the standard condition (13).

D. Related Results

We note that Theorem 8 is in the spirit of the OMP recovery condition derived recently in [15, Thm. 1]. While the condition in [15, Thm. 1] imposes a constraint on the decay between consecutive magnitudes

in \mathbf{x} (ordered in decaying fashion) and requires perfectly sparse vectors that satisfy $\|\mathbf{x}\|_0 < 1/\mu_a$. In contrast, Theorem 8 makes explicit use of the δ -density and ensures that OMP identifies the largest t_{\max} entries, irrespective of whether the signal \mathbf{x} to be recovered is perfectly sparse or not. A similar recovery result for OMP to recover signals with fast decaying entries has been developed in [16]; this result, however, relies on the restricted isometry property (RIP), which is a common way of characterizing measurement matrices in the field of compressive sensing [17] but hard to compute in practice.

V. CONCLUSIONS

We have developed a novel kernel result and uncertainty relation for signals with low density. Our results generalize existing results for sparse signals to non-sparse signals with sufficiently low density. Furthermore, we have shown that OMP recovers signals with low density and fast-decaying magnitudes. Our results highlight that an appropriately chosen *density measure* is key for establishing kernel results, uncertainty relations, or recovery guarantees, in contrast to results that solely rely on signal sparsity.

There are many avenues for future work. By modifying the proof of Theorem 8, one can obtain a recovery condition in presence of bounded measurement noise $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ with $\|\mathbf{n}\|_2 \leq \varepsilon$ using techniques developed in [4]. In addition, our proofs can easily be extended to block-sparse signals [18]. An analysis of alternative density measures that behave similarly to our δ -density, including $\gamma(\mathbf{x}) = \|\mathbf{x}\|_2^2 / \|\mathbf{x}\|_\infty^2$ and $\sigma(\mathbf{x}) = \|\mathbf{x}\|_1^2 / \|\mathbf{x}\|_2^2$ (which was used in, e.g., [19]–[21]), is left for future work. Another interesting direction is the analysis of optimization-based recovery of signals with low density.

ACKNOWLEDGMENTS

The author thanks Tom Goldstein and Helmut Bölcskei for insightful discussions, and Ramina Ghods and Charles Jeon for comments on an early version of the paper.

APPENDIX A

BASIC PROPERTIES

A. Proof of Lemma 1

From Definition 1, we have

$$\delta(\mathbf{x}) = \frac{\sum_{i \in \mathcal{X}} |x_i|}{\max_k |x_k|} = \sum_{i \in \mathcal{X}} \frac{|x_i|}{|x_{\hat{k}}|} \leq |\mathcal{X}| = \|\mathbf{x}\|_0, \quad (14)$$

where \mathcal{X} is the support set of the signal \mathbf{x} and \hat{k} is the index associated to (one of) the largest coefficient(s) of \mathbf{x} in terms of magnitude. The inequality in (14) follows from the fact that $|x_i|/|x_{\hat{k}}| \leq 1$, for $i \in \mathcal{X}$.

Equality holds if and only if $|x_i|/|x_{\hat{k}}| = 1$, $i \in \mathcal{X}$, which implies that only signals whose nonzero entries have constant modulus satisfy $\delta(\mathbf{x}) = \|\mathbf{x}\|_0$.

B. Proof of Lemma 3

The proof follows from a counterexample. Consider two signals $\mathbf{x} \in \mathbb{R}^{N_a}$ and $\mathbf{z} \in \mathbb{R}^{N_a}$ with entries $x_i = -\alpha^{i-1}$ and $z_i = \alpha^{i-1} + \varepsilon$ for $i = 1, \dots, N_a$ with $0 < \alpha < 1$ and $\varepsilon \geq 0$. Both vectors have the following δ -densities:

$$\delta(\mathbf{x}) = \frac{1 - \alpha^{N_a}}{1 - \alpha} \quad \text{and} \quad \delta(\mathbf{z}) = \frac{1}{1 + \varepsilon} \left(\frac{1 - \alpha^{N_a}}{1 - \alpha} + N_a \varepsilon \right),$$

which can be made small by reducing α . However, the sum of both signals has constant-modulus (with magnitude ε), i.e.,

$$\delta(\mathbf{x} + \mathbf{z}) = \delta(\varepsilon \mathbf{1}_{N_a \times 1}) = N_a.$$

Hence, as $\delta(\mathbf{x} + \mathbf{z}) \geq \delta(\mathbf{x}) + \delta(\mathbf{z})$, these signals violate the triangle inequality for sufficiently small values of α and ε , and a suitably large ambient dimension N_a .

APPENDIX B

PROOF OF KERNEL RESULT AND UNCERTAINTY RELATION

A. Proof of Lemma 4

We start with the lower bound in (3). Since \mathbf{A} is a dictionary with $\mathbf{a}_i^H \mathbf{a}_i = 1$, $\forall i$, we obtain the following lower bound:

$$\|\mathbf{A}^H \mathbf{A} \mathbf{x}\|_\infty = \max_i \left| \mathbf{a}_i^H \mathbf{a}_i x_i + \sum_{j, i \neq j} \mathbf{a}_i^H \mathbf{a}_j x_j \right| \geq \max_i \left\{ |x_i| - \sum_{j, i \neq j} |\mathbf{a}_i^H \mathbf{a}_j| |x_j| \right\},$$

where the last step follows from the reverse and regular triangle inequalities, respectively. Using the definition of the coherence μ_a , we obtain

$$\begin{aligned} \|\mathbf{A}^H \mathbf{A} \mathbf{x}\|_\infty &\geq \max_i \left\{ |x_i| - \sum_{j, i \neq j} \mu_a |x_j| \right\} \\ &= \max_i \left\{ |x_i| (1 + \mu_a) - \mu_a \sum_j |x_j| \right\} \\ &= \|\mathbf{x}\|_\infty (1 + \mu_a) - \mu_a \|\mathbf{x}\|_1. \end{aligned} \tag{15}$$

By excluding the case $\|\mathbf{x}\|_\infty = 0$ (implying $\mathbf{x} \neq \mathbf{0}_{N_a \times 1}$), we finally get the lower bound in (3). The upper bound in (3) is obtained similarly to (15) and is given by

$$\begin{aligned} \|\mathbf{A}^H \mathbf{A} \mathbf{x}\|_\infty &\leq \max_i \left\{ |x_i| + \sum_{j, i \neq j} |\mathbf{a}_i^H \mathbf{a}_j| |x_j| \right\} \\ &\leq \|\mathbf{x}\|_\infty (1 - \mu_a) + \mu_a \|\mathbf{x}\|_1. \end{aligned} \quad (16)$$

By excluding $\|\mathbf{x}\|_\infty = 0$, we obtain the upper bound in (3).

B. Proof of Theorem 6

To prove Theorem 6, we will use the following result:

$$\|\mathbf{A}^H \mathbf{B} \mathbf{z}\|_\infty = \max_i |\mathbf{a}_i^H \mathbf{B} \mathbf{z}| = \max_i \left| \sum_k \mathbf{a}_i^H \mathbf{b}_k z_k \right| \leq \max_i \sum_k \mu_m |z_k| = \mu_m \|\mathbf{z}\|_1. \quad (17)$$

Analogously, we will use $\|\mathbf{B}^H \mathbf{A} \mathbf{x}\|_\infty \leq \mu_m \|\mathbf{x}\|_1$. Now, if $\mathbf{A} \mathbf{x} = \mathbf{B} \mathbf{z}$, then we have $\mathbf{A}^H \mathbf{A} \mathbf{x} = \mathbf{A}^H \mathbf{B} \mathbf{z}$ and also $\|\mathbf{A}^H \mathbf{A} \mathbf{x}\|_\infty = \|\mathbf{A}^H \mathbf{B} \mathbf{z}\|_\infty$. Using the LHS of (3) and (17), we can bound $\|\mathbf{A}^H \mathbf{A} \mathbf{x}\|_\infty$ from below and $\|\mathbf{A}^H \mathbf{B} \mathbf{z}\|_\infty$ from above, respectively, and obtain

$$[1 - \mu_a(\delta(\mathbf{x}) - 1)]^+ \|\mathbf{x}\|_\infty \leq \mu_m \|\mathbf{z}\|_1, \quad (18)$$

where we take the non-negative part in the LHS because norms are non-zero. Since $\mathbf{A} \mathbf{x} = \mathbf{B} \mathbf{z}$, we also have $\|\mathbf{B}^H \mathbf{A} \mathbf{x}\|_\infty = \|\mathbf{B}^H \mathbf{B} \mathbf{z}\|_\infty$. Using similar steps as above, we obtain

$$[1 - \mu_b(\delta(\mathbf{z}) - 1)]^+ \|\mathbf{z}\|_\infty \leq \mu_m \|\mathbf{x}\|_1. \quad (19)$$

Multiplying (18) with (19) and dividing both sides by $\|\mathbf{x}\|_\infty \|\mathbf{z}\|_\infty$ yields the uncertainty relation (5).

APPENDIX C

PROOF OF THEOREM 8

A. Preliminaries

Without loss of generality, we sort the coefficients of \mathbf{x} in descending order of their magnitudes $|x_i|$, and also sort the corresponding atoms in \mathbf{A} ; ties are broken arbitrarily. In what follows, we consider the resulting “sorted” system $\mathbf{y} = \mathbf{A} \mathbf{x}$.

For OMP to recover the i -th largest atom in iteration i , we need the following condition [4]

$$\max_{i \in \mathcal{M}_S} |\mathbf{a}_i^H \mathbf{R}_S \mathbf{y}| > \max_{k \in \mathcal{M}_S^c \setminus \mathcal{S}} |\mathbf{a}_k^H \mathbf{R}_S \mathbf{y}|$$

to hold. Here, the set

$$\mathcal{M}_S = \{k : \max_{i \in \mathcal{S}^c} |x_i| = |x_k|, \forall k\}$$

contains all indices corresponding to entries having the same (and largest magnitude), excluding the already taken elements in \mathcal{S} , and $\mathbf{R}_S = \mathbf{I}_M - \mathbf{A}_S \mathbf{A}_S^\dagger$ is the projector onto the orthogonal complement of the columns spanned by the atoms in the set \mathcal{S} previously selected by OMP. Since $\mathbf{y} = \mathbf{a}_i x_i + \mathbf{A}_S \mathbf{x}_S + \mathbf{A}_{\mathcal{R} \setminus i} \mathbf{x}_{\mathcal{R} \setminus i}$, where $\mathcal{R} = \{1, \dots, N_a\} \setminus \mathcal{S}$, we get the following equivalent condition:

$$\begin{aligned} \max_{i \in \mathcal{M}_S} |\mathbf{a}_i^H \mathbf{R}_S (\mathbf{a}_i x_i + \mathbf{A}_S \mathbf{x}_S + \mathbf{A}_{\mathcal{R} \setminus i} \mathbf{x}_{\mathcal{R} \setminus i})| &> \\ \max_{k \in \mathcal{M}_S^c \setminus \mathcal{S}} |\mathbf{a}_k^H \mathbf{R}_S (\mathbf{a}_i x_i + \mathbf{A}_S \mathbf{x}_S + \mathbf{A}_{\mathcal{R} \setminus i} \mathbf{x}_{\mathcal{R} \setminus i})|. \end{aligned} \quad (20)$$

The fact that $\mathbf{R}_S \mathbf{A}_S = \mathbf{0}_{M \times |\mathcal{S}|}$ allows us to simplify (20) to

$$\max_{i \in \mathcal{M}_S} |\mathbf{a}_i^H \mathbf{R}_S (\mathbf{a}_i x_i + \mathbf{A}_{\mathcal{R} \setminus i} \mathbf{x}_{\mathcal{R} \setminus i})| > \max_{k \in \mathcal{M}_S^c \setminus \mathcal{S}} |\mathbf{a}_k^H \mathbf{R}_S \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}}|. \quad (21)$$

In order to arrive at a sufficient condition for OMP to select the correct atom in iteration i , we now individually lower and upper-bound the LHS and RHS of (21), respectively.

B. Lower Bound on the LHS of (21)

We start with the expansion

$$\begin{aligned} \max_{i \in \mathcal{M}_S} |\mathbf{a}_i^H (\mathbf{I}_M - \mathbf{A}_S \mathbf{A}_S^\dagger) (\mathbf{a}_i x_i + \mathbf{A}_{\mathcal{R} \setminus i} \mathbf{x}_{\mathcal{R} \setminus i})| &= \\ \max_{i \in \mathcal{M}_S} |x_i + \mathbf{a}_i^H \mathbf{A}_{\mathcal{R} \setminus i} \mathbf{x}_{\mathcal{R} \setminus i} - \mathbf{a}_i^H \mathbf{A}_S \mathbf{A}_S^\dagger (\mathbf{a}_i x_i + \mathbf{A}_{\mathcal{R} \setminus i} \mathbf{x}_{\mathcal{R} \setminus i})| \end{aligned}$$

and apply the reverse triangle inequality to obtain the following lower bound:

$$\begin{aligned} \max_{i \in \mathcal{M}_S} |x_i| - |\mathbf{a}_i^H \mathbf{A}_{\mathcal{R} \setminus i} \mathbf{x}_{\mathcal{R} \setminus i}| - |\mathbf{a}_i^H \mathbf{A}_S \mathbf{A}_S^\dagger (\mathbf{a}_i x_i + \mathbf{A}_{\mathcal{R} \setminus i} \mathbf{x}_{\mathcal{R} \setminus i})| \\ \geq \max_{i \in \mathcal{M}_S} |x_i| (1 + \mu_a) - \sum_{k \in \mathcal{R}} \mu_a |x_k| - \sum_{k \in \mathcal{R}} |\mathbf{a}_i^H \mathbf{A}_S \mathbf{A}_S^\dagger \mathbf{a}_k| |x_k|. \end{aligned}$$

We now bound the last term in the above result as follows:

$$\begin{aligned} |\mathbf{a}_i^H \mathbf{A}_S \mathbf{A}_S^\dagger \mathbf{a}_k| &\leq \|\mathbf{A}_S^H \mathbf{a}_i\|_2 \|(\mathbf{A}_S^H \mathbf{A}_S)^{-1} \mathbf{A}_S^H \mathbf{a}_k\|_2 \\ &\leq \frac{\|\mathbf{A}_S^H \mathbf{a}_i\|_2 \|\mathbf{A}_S^H \mathbf{a}_k\|_2}{[1 - \mu_a(|\mathcal{S}| - 1)]^+} \leq \frac{\mu_a^2 |\mathcal{S}|}{[1 - \mu_a(|\mathcal{S}| - 1)]^+}. \end{aligned} \quad (22)$$

where we used the Cauchy-Schwarz inequality, Geršgorin's disc theorem [22, Thm. 6.1.1], and the definition of the mutual coherence μ_m . We note that (22) requires

$$|\mathcal{S}| < 1/\mu_a + 1, \quad (23)$$

which must hold for any set size picked by OMP; as a result, we get condition (8). By combining the above bounds, we arrive at

$$\max_{i \in \mathcal{M}_S} |\mathbf{a}_i^H \mathbf{R}_S \mathbf{y}| \geq \max_{i \in \mathcal{M}_S} |x_i| (1 + \mu_a) - \frac{\mu_a + \mu_a^2}{[1 - \mu_a(|\mathcal{S}| - 1)]^+} \sum_{k \in \mathcal{R}} |x_k|. \quad (24)$$

C. Upper Bound on the RHS of (21)

We now bound the RHS in (21) from above. To this end, we first expand

$$|\mathbf{a}_k^H \mathbf{R}_S \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}}| = |\mathbf{a}_k^H (\mathbf{I}_M - \mathbf{A}_S \mathbf{A}_S^\dagger) \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}}| = |\mathbf{a}_k^H \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}} - \mathbf{a}_k^H \mathbf{A}_S \mathbf{A}_S^\dagger \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}}|$$

and use the triangle inequality to get

$$|\mathbf{a}_k^H \mathbf{R}_S \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}}| \leq |\mathbf{a}_k^H \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}}| + |\mathbf{a}_k^H \mathbf{A}_S \mathbf{A}_S^\dagger \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}}|.$$

The first RHS term can be bounded as follows:

$$|\mathbf{a}_k^H \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}}| \leq \sum_{k \in \mathcal{R}} \mu_a |x_k|.$$

The second RHS term can be bounded using the same approach used in (22) to get

$$|\mathbf{a}_k^H \mathbf{A}_S \mathbf{A}_S^\dagger \mathbf{A}_{\mathcal{R}} \mathbf{x}_{\mathcal{R}}| \leq \frac{\mu_a^2 |\mathcal{S}|}{[1 - \mu_a(|\mathcal{S}| - 1)]^+} \sum_{k \in \mathcal{R}} |x_k|$$

Combining the above results leads to the following bound:

$$\max_{k \in \mathcal{M}_S^c \setminus \mathcal{S}} |\mathbf{a}_k^H \mathbf{R}_S \mathbf{y}| \leq \frac{\mu_a + \mu_a^2}{[1 - \mu_a(|\mathcal{S}| - 1)]^+} \sum_{k \in \mathcal{R}} |x_k|. \quad (25)$$

D. Combining Both Bounds

By combining (24) and (25) and using simple algebraic manipulations, we get the following sufficient condition for OMP to select (one of) the largest column(s) of \mathbf{A} :

$$\max_{i \in \mathcal{M}_S} |x_i| > \frac{2\mu_a}{[1 - \mu_a(|\mathcal{S}| - 1)]^+} \sum_{k \in \mathcal{R}} |x_k|.$$

Assuming (23) is satisfied, $|x_i| \neq 0$, and $|\mathcal{S}| = t - 1$, we finally get the following condition:

$$\frac{1}{2} (1/\mu_a + 1 - (t - 1)) > \sum_{k \in \mathcal{R}} \frac{|x_k|}{\max_{i \in \mathcal{M}_s} |x_i|} = \delta(\mathbf{x}_{\mathcal{R}}),$$

which ensures that OMP selects an atom i associated to the largest remaining coefficient x_i (in magnitude).

REFERENCES

- [1] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Th.*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.
- [2] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Th.*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
- [3] —, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Th.*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [4] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Th.*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
- [5] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.
- [6] M. Elad, "Sparse representations are most likely to be the sparsest possible," *EURASIP J. App. Sig. Proc.*, vol. 7, pp. 1–12, 2006.
- [7] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Th.*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.
- [8] P. Kuppinger, G. Durisi, and H. Bölcskei, "Uncertainty relations and sparse signal recovery for pairs of general signal sets," *IEEE Trans. Inf. Th.*, vol. 58, no. 1, pp. 263–277, Jan. 2012.
- [9] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Th.*, vol. 58, no. 5, pp. 3115–3130, May 2012.
- [10] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Sig. Proc.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.
- [11] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st ed. Springer, Aug. 2010.
- [12] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, Jun. 1989.
- [13] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. of 27th Asilomar Conf. on Signals, Systems, and Comput.*, Nov. 1993, pp. 40–44.
- [14] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.
- [15] C. Herzet, A. Drémeau, and C. Soussen, "Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay," *arXiv*, vol. 1401.7533v2, Aug. 2014.
- [16] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Th.*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.
- [17] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [18] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Sig. Proc.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
- [19] G. Tang and A. Nehorai, "Performance analysis of sparse recovery based on constrained minimal singular values," *IEEE Trans. Sig. Proc.*, vol. 59, no. 12, pp. 5734–5745, Dec. 2011.
- [20] J. Andersson and J.-O. Strömberg, "On the theorem of uniform recovery of random sampling matrices," *IEEE Trans. Inf. Th.*, vol. 60, no. 3, pp. 1700–1710, Mar. 2014.
- [21] E. Esser, Y. Lou, and J. Xin, "A method for finding structured sparse solutions to non-negative least squares problems with applications," *arXiv*, vol. 1301.0413v1, Jan. 2013.
- [22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge Press, 1985.